Speed and impact of team science during urgent societal events

Nicholas A. Coles¹*, Joao Francisco Goes Braga Takayanagi¹, Stephen M. Fiore², Lingfei Wu³*

¹ Department of Psychology, University of Florida, Gainesville, FL, USA

² Institute of Simulation and Training, University of Central Florida, Orlando, FL, USA

³ School of Computing and Information, University of Pittsburgh, Pittsburgh, PA, USA

^{*} Corresponding authors: Nicholas A. Coles (ncoles@ufl.edu) and Lingfei Wu (liw105@pitt.edu)

Abstract

Urgent societal events demand scientific responses that are both rapid and impactful. Through an adversarial collaboration, we connected bibliometric databases to evaluate the speed and impact of over 2 million scientific publications in the three years following 48 urgent societal events. A pilot analysis of three cases — the 2022 release of ChatGPT, the 2019 COVID-19 pandemic, and the 2001 World Trade Center attacks — yielded unexpected patterns: larger teams were not only more impactful but also quicker to publish. More precisely, increases in team size were associated with (a) initial increases, but eventual diminishing returns in academic citations, (b) curvilinear returns in news and policy document citations, and (c) curvilinear returns in terms of how quickly papers were published. In other words, there are points where further increases in team sizes are either marginally helpful (diminishing returns) or counterproductive (curvilinear returns). To evaluate robustness, we pre-registered a broader test covering 45 additional events spanning two decades.

Keywords: team science, urgent societal events, speed, impact

Introduction

In the first quarter of the 21st century alone, society encountered numerous events that demanded quick, urgent, and widespread action. This includes (a) the sudden public release and uptake of disruptive AI technologies, (b) a global pandemic, and (c) some of the deadliest terrorist attacks in modern history. *Urgent societal events*, such as these, represent historically pivotal moments – moments that often signal an impending period of transformative social, political, and/or technological change. Despite the potential significance of science in these moments, a fundamental question remains unresolved: How should the ecosystem of science best organize its response? For example, should researchers, funders, and stakeholders prioritize larger, more collectivistic research projects – or smaller, more individualistic efforts?

Previous scholarship on the benefits and drawbacks of large teams in science focuses on how scientists operate in ordinary conditions – not historical moments where society has a sudden, intense, and urgent need for scientific insights. Notably, team dynamics that work well in ordinary conditions do not always generalize in moments of crisis^{1,2}. Unfortunately, our understanding of team science in these moments is largely based on anecdotes and small samples^{3,4}. In the present work, we more comprehensively take stock of scientists' responses to urgent societal events by connecting bibliometric databases containing over 250 million scientific publications, 2.6 billion citations in scholarly works, 25 million mentions in news outlets, and 25 million mentions in public policy documents. In doing so, we examine the success of scientific responses in terms of (a) how quickly they publish their insights and (b) how much attention these insights receive in science, mass media, and public policy. We first explored these patterns in a pilot investigation of three high-profile events. We then committed to a Registered Report to evaluate the robustness and generalizability of these results across 45 additional urgent events spanning over two decades.

We began our adversarial collaboration¹ with an informal consensus: during urgent societal events, larger teams will be more impactful *but slower*. There are several reasons to expect larger teams to be more impactful. Large-scale collaboration (a) allows role specialization¹¹, (b) increases total effort dedicated to a problem⁸, and (c) creates conditions for facilitating "recombinant growth"^{12,13}, uncovering the "wisdom of the crowd"^{14,15}, manifesting "collective intelligence"¹⁶, and leveraging social channels to spread impact¹⁷. In ordinary conditions, bibliometric investigations paint a clear picture: larger teams receive more citations in scholarly articles^{9,17–20}, patents⁹, code repositories⁹, news outputs²¹, and policy documents²¹. This impact, however, is often suggested to come at a cost that may be particularly noteworthy during urgent societal events: speed. Speed is rarely (if ever) considered in bibliometric investigations. However, it takes time to assemble a large team¹⁶, negotiate responsibilities¹⁴, navigate different institutional policies¹⁵, coordinate the work²², track contributions¹⁷, and manage disagreements¹⁷. Working with a small team may be quicker, with more streamlined coordination, fewer bureaucratic constraints, and greater decision-making agility.

A pilot investigation yielded surprising results: big teams were not only more impactful, but also slightly *quicker* to publish their insights. This investigation involved research articles containing keywords linked to three events: (1) the initial release of the generative AI tool, ChatGPT, (2) the COVID-19 pandemic, and

¹ This work is the result of an *adversarial collaboration*: a procedure in which scholars with diverging perspectives work together to accelerate progress on an outstanding scientific issue^{5–7}. Collectively, our collaboration involves researchers who have historically advocated for (a) relatively *big* teams in science⁸, (b) relatively *small* teams in science⁹, and (c) more neutrally, team science efforts designed to bridge

⁽b) relatively *small* teams in science⁹, and (c) more neutrally, team science efforts designed to bridge differing perspectives on complex issues¹⁰.

(3) the 2001 World Trade Center attacks. (See *Methods* for more information.) We examined the speed of scientists' published responses by leveraging a unique feature of urgent societal events: there is an approximate date when the event entered mainstream public awareness (e.g., ChatGPT was publicly released November 2022). Response speed was measured as the number of days between the onset of each event and the publication date of each article mentioning associated keywords (e.g., keyword "ChatGPT"). In addition to their speed, we examined the impact of these articles across several domains of society. Particularly during urgent societal events, the ecosystem of science needs to consider both the academic and societal impact of their work. Thus, we not only measured the total number of citations received from other scholarly articles, but also news media and policy documents.

For links between team size, speed, and impact, we evaluated three rates of return (see *Methods*). First, we estimated the extent to which increases in team sizes are associated with *constant* (linear) rates of returns in speed and impact. Second, we evaluated *diminishing returns*²³: a "plateau effect" wherein scientists must make increasingly large additions to their teams to increase the speed and impact of their work. Third, we evaluated *curvilinear returns*: a "Goldilocks effect" wherein increases in team size lead to improved performance up to a point, after which further growth becomes detrimental.

In our pilot analysis, the relationship between team size and scholarly citations was best captured by a model indicating a "plateau effect" or diminishing returns. For all other outcomes, models representing curvilinear "Goldilocks effects" achieved best fit. While increases in team size initially benefitted non-scholarly impact and speed, performance generally declined beyond a threshold. These estimated points of decline were 75 co-authors for news citations (z = 13.33), 85 for policy citations (z = 15.28), and 49 for speed (z = 8.42). As implied by z-scores, teams rarely get this large; 75% of papers had 5 or fewer co-authors, 97% had 15 or fewer, and 99% had 25 or fewer. In other words, in terms of fast and impactful responses to urgent societal events, the optimum team size is rarely reached or exceeded (see Supplementary Figure 1)²¹.

These results – if reliable – have implications for both theory and practice. Theoretically, they help uncover the functional dynamics of a socially unique response to urgent societal events: mass collaboration^{24–26}. Practically, such findings may also inform evidence-based scientific responses to *future* urgent societal events – such as an intensified "Al race"^{27,28}, a potential H5N1 avian influenza virus outbreak²⁹, or efforts to radically shift political, economic, and social systems³⁰. Indeed, if building up big teams (at least up to a certain size) produces faster and more impactful papers, the scientific enterprise would have to undergo substantial transformation to better align norms, incentives, funding models, and infrastructure to support larger teams^{8,22,31}.

In the present work, we sought to evaluate the replicability, generalizability, and robustness of these pilot study insights (see Table 1). First, we committed to a Registered Report format, seeking to improve the trustworthiness of our investigation by pre-registering our methodological approach, deliberating details *before* final data collection commences, and committing to publishing the results regardless of the findings³². Second, we evaluated the reliability of these patterns across 45 additional urgent societal events from the past two decades³³. In other words, we evaluate the extent to which context matters – a key factor to consider when making future recommendations for novel urgent societal events. Last, we pre-registered secondary analyses designed to probe the robustness of our findings.

Methods

Ethics information

Records were sourced from OpenAlex: an openly available catalogue of scientific articles³⁴. To evaluate cumulative impact, we examined the total number of citations received from other scholarly publications indexed by OpenAlex – as well as media and policy document citations indexed by a proprietary service, Altmetric (http://www.altmetric.com/). Because we relied exclusively on secondary data, IRB approval was not required.

Sampling plan

Primarily, we designed our sampling plan to provide > 95% power to replicate the best-fitting model in our pilot analysis (see Table 1). To estimate power, we performed Monte Carlo simulations (500 iterations). Team size was simulated by randomly sampling from the distribution observed in pilot data. Predicted speed and impact values were simulated from mixed-effect regression models fitted on pilot data. Power for detecting hypothesized diminishing returns (in scholarly citations) and curvilinear returns (in news citations, policy citations, and speed) was operationalized as the proportion of Monte Carlo simulation iterations wherein the pre-registered regression model yielded lower squared error than a comparison model containing only a linear term. Squared error distributions were compared using Wilcoxon signed-rank tests. Results indicated that 100,000 papers would provide 95% power to detect all hypothesized relationships in Table 1. For more information, see *SI*.

Secondarily, we designed our sampling plan to estimate the extent to which observed patterns generalize to a broader set of urgent societal events. To ensure that estimates of between-event heterogeneity were reliable, we followed previous recommendations and sought to sample between 15 and 50 events³⁵.

Identifying urgent societal events

Our operationalization of "urgent societal events" focused on two inclusion criteria. First, the event had to be generally recognized as "urgent" by members of society. Second, the event had to be associated with a large scientific response – both in terms of the absolute number of publications linked to each topic as well as its growth.

To identify events widely recognized as "urgent" by members of society, we used the large language model (LLM), Claude 3.7 Sonnet³⁶. LLMs, like Claude 3.7 Sonnet, are trained on massive text corpora – including posts, news articles, and papers written about urgent societal events. In addition to their ability to detect sentiments and attitudes³⁷, LLMs are increasingly being used for event and information extraction³⁸. We leveraged this approach to compile an initial list of ≈100 urgent societal events, passing the following prompt to Claude 3.7 (paraphrased for length):

"...Identify 100 urgent societal events: historical moments in the 21st century where society had a sudden, intense, and urgent need for scientific insights...mark the moment in which the events became a global concern...identify 3 keywords that uniquely identify each event..."

Two independent human raters reviewed the list of events, dates, and keywords generated by the LLM. In all cases, they agreed with the output.

Next, we used OpenAlex to ensure that each event was associated with sizable growth in scientists' responses. For example, we identified N = 57,010 publications that mention the phrase "Zika virus." In the three years prior to the World Health Organisation declaring a Public Health Emergency of

International Concern (February 2016³⁹), only 374 papers contained this keyword. Over the next three years, 19,610 more publications would emerge – a 52-fold increase. In other words, in addition to being widely recognized as an "urgent societal event", the 2015 Zika virus epidemic is associated with a large scientific response.

To be included in our analysis, events must also have had at least 500 papers that mentioned an associated keyword (according to an initial OpenAlex search). Furthermore, the population of papers mentioning the keyword must have experienced at least 2-fold growth in the three years following the urgent societal event. Forty-five urgent societal events met these inclusion criteria, leaving us with an estimated 336,965 previously-unexamined papers. This was used as an *initial* estimate used to establish confidence in statistical power; to ensure the most accurate and up-to-date results, the final set of records was not downloaded and processed until after initial Stage 1 acceptance.

Analyses focused on works published within three years following each urgent societal event. Works were removed from analyses if OpenAlex indicated that it was (a) not an article (e.g., a pre-print or book), (b) missing a persistent identifier (e.g., a Digital Object Identifier or a PubMed ID), or (c) completely missing authorship meta-data. Works were also removed if their title contained a phrase that indicated it may have been misclassified as a research article: "transactions of", "transactions on", "proceeding of", "conference", "annual meeting", "symposium", and "table of contents"²¹.

Scholarly citation counts were extracted from OpenAlex. Using persistent identifiers (Digital Object Identifiers and PubMed IDs), Altmetric Explorer was used to extract estimates of the total citations in mainstream news outlets (e.g., the *New York Times*) and blog posts (e.g., *Behind the Paper*), which we subsequently summed. Altmetric Explorer was also used to extract estimates of the total citations in public policy documents (e.g., from the *Publications Office of the European Union*). To reduce sensitivity to outliers and misspecified distributions, speed and impact indices were converted to percentiles, calculated separately for each event and coded so that higher values represented greater impact and faster publication.

Analysis plan

As shown in Table 1, we used weighted mixed-effect regression to separately model the relationship between team size and four outcomes: publication speed, and citations in scholarly articles, the news, and public policy documents. We used mixed-effect regression in order to estimate both fixed effects (i.e., the overall effect across all events) and random effects (i.e., variation in the effect across events). This allowed us not only to test the replicability of our pilot findings, but also to estimate the reliability of these patterns across a broader set of urgent societal events. To ensure that relatively uncommon large team efforts (e.g., 20 co-authors) received the same priority as relatively common small team efforts (e.g., 2 co-authors) during model fitting, we applied inverse frequency weights. This helped ensure that our models did not overfit on papers by small teams and underfit on papers by large teams.

We first tested the replicability of our pilot findings by separately modeling linear, logarithmic, and quadratic relationships between team size and each outcome. To accommodate the fact that papers published on similar topics (e.g., "Zika virus") are non-independent, we included random intercepts for each sampled urgent societal event ($\beta_{oe} = \beta_o + u_{oe}$). When comparing models, we pre-registered that a model would be favored if the Wilcoxon signed-rank test indicated a significantly lower distribution of root squared error (p < .05). As an additional basis for model comparison, we also estimated and reported Bayesian Information Criteria (BIC). After evaluating the replicability of our pre-registered model, we added random slopes (e.g., $\beta_{1e} = \beta_1 + u_{1e}$) in order to quantify the extent to which the team-size effect (β_1) varies across events (u_{1e}).

<u>Secondary analyses: Evaluating potential confounds</u>

Although questions about underlying mechanisms were beyond the scope of our investigation, we identified three sets of analyses we believed would productively contribute to future discussions. Unlike our primary analyses, these secondary analyses were specified *without* prior pilot testing.

One possibility is that more successful fields, subfields, or topics tend to attract larger teams – but that these large teams don't have a performance advantage within those domains. In other words, large teams may be drawn to high-impact areas – but possess no unique ability to produce high-impact insights in those areas. To evaluate this possibility, we re-ran the primary analyses using three standardization approaches: by the paper's primary field, subfield, or topic (as identified by OpenAlex). For example, consider field-level standardization. For each event, we (a) grouped papers by their primary field, (b) calculated impact and speed percentiles within each field, and (c) re-ran our primary analyses using these standardized values. Conceptually, these models estimate the relationship between team size and speed/impact while accounting for potentially confounding between-field [between-subfield, between-topic] differences in the overall speed and impact of teams.

A second possibility is that the relationships between team size, speed, and impact of teams are driven by prestige. Prestigious institutions may hold economic, social, and cultural capital that amplifies their research impact — capital that may also uniquely position them to lead big team science initiatives. To evaluate this possibility, we re-ran primary analyses with a fourth standardization approach: standardizing speed and impact metrics by the institution of each paper's first author. We ran an additional analysis that standardized by the paper's last author, a second common spot for the lead author. Conceptually, these models estimate the relationship between team size and outcomes while controlling for potentially confounding institutional differences in typical team size and impact.

A third possibility is that larger teams generate more impact by publishing more quickly. Research on the 'priority rule' suggests that stakeholders favor insights from teams that publish first^{40,41}. To evaluate this, we re-ran our primary analyses with one modification: speed was included as a mean-centered fixed effect. Conceptually, these models estimate the relationship between team size and impact while controlling for potential confounding with speed.

References

- 1. Hannah, S. T., Uhl-Bien, M., Avolio, B. J. & Cavarretta, F. L. A framework for examining leadership in extreme contexts. *Leadersh. Q.* **20**, 897–919 (2009).
- Ratcharak, P. Leadership dynamics in health care crises: The impact of initiating structure and consideration behaviors on safety climate in public hospitals. *Health Care Manage. Rev.* 50, 221–231 (2025).
- 3. Buisse, E., Bubbers, J., Sommers, H., Stavrou, V. & Denis, M. Protecting Science in Times of Crisis. *Int. Sci. Counc.* (2024) doi:10.24948/2024.01.
- 4. Wolbers, J., Kuipers, S. & Boin, A. A systematic review of 20 years of crisis and disaster research:

 Trends and progress. *Risk Hazards Crisis Public Policy* **12**, 374–392 (2021).
- Coles, N. A. et al. A multi-lab test of the facial feedback hypothesis by the Many Smiles
 Collaboration. Nat. Hum. Behav. 6, 1731–1742 (2022).
- Nature. Make science more collegial: why the time for 'adversarial collaboration' has come. *Nature* 641, 281–282 (2025).
- 7. Ferrante, O. *et al.* Adversarial testing of global neuronal workspace and integrated information theories of consciousness. *Nature* **642**, 133–142 (2025).
- 8. Coles, N. A., Hamlin, J. K., Sullivan, L. L., Parker, T. H. & Altschul, D. Build up big-team science. *Nature* **601**, 505–507 (2022).
- 9. Wu, L., Wang, D. & Evans, J. A. Large teams develop and small teams disrupt science and technology.

 Nature 566, 378–382 (2019).
- Hall, K. L. *et al.* The science of team science: A review of the empirical evidence and research gaps on collaboration in science. *Am. Psychol.* **73**, 532–548 (2018).
- 11. Tiokhin, L., Panchanathan, K., Smaldino, P. E. & Lakens, D. Shifting the Level of Selection in Science.

 Perspect. Psychol. Sci. 19, 908–920 (2024).

- 12. Weitzman, M. L. Recombinant Growth. Q. J. Econ. 113, 331–360 (1998).
- 13. Uzzi, B., Mukherjee, S., Stringer, M. & Jones, B. Atypical Combinations and Scientific Impact. *Science* **342**, 468–472 (2013).
- 14. Simoiu, C., Sumanth, C., Mysore, A. & Goel, S. Studying the "Wisdom of Crowds" at Scale. *Proc. AAAI Conf. Hum. Comput. Crowdsourcing* **7**, 171–179 (2019).
- 15. Surowiecki, J. *The Wisdom of Crowds: Why the Many Are Smarter than the Few and How Collective Wisdom Shapes Business, Economies, Societies, and Nations*. xxi, 296 (Doubleday & Co, New York, NY, US, 2004).
- 16. Mulgan, G. *Big Mind : How Collective Intelligence Can Change Our World*. (Princeton University Press, Princeton, NJ, 2017).
- 17. Coles, N. A. The Prevalence, Impact, and Nature of Big Team Science. Preprint at https://doi.org/10.31234/osf.io/q68yv (2024).
- 18. Hsiehchen, D., Espinoza, M. & Hsieh, A. Multinational teams and diseconomies of scale in collaborative research. *Sci. Adv.* **1**, e1500211 (2015).
- 19. Larivière, V., Gingras, Y., Sugimoto, C. R. & Tsou, A. Team size matters: Collaboration and scientific impact since 1900. *J. Assoc. Inf. Sci. Technol.* **66**, 1323–1332 (2015).
- 20. Jones, B. F. The Rise of Research Teams: Benefits and Costs in Economics. *J. Econ. Perspect.* **35**, 191–216 (2021).
- 21. Coles, N. A., Takayanagi, J. F. G. B., Grant, G. L. & Basnight-Brown, D. The Rise, Impact, and Imbalances of Big Team Psychology. Preprint at https://doi.org/10.31234/osf.io/b3mq6_v1 (2025).
- 22. Baumgartner, H. A. *et al.* How to build up big team science: a practical guide for large-scale collaborations. *R. Soc. Open Sci.* **10**, 230235 (2023).
- 23. Mao, A., Mason, W., Suri, S. & Watts, D. J. An Experimental Study of Team Size and Performance on a Complex Task. *PLOS ONE* **11**, e0153048 (2016).

- 24. Pan, S., Pan, G. & Leidner, D. Crisis Response Information Networks. *J. Assoc. Inf. Syst.* **13**, 31–56 (2012).
- 25. Drury, J., Cocking, C. & Reicher, S. Everyone for themselves? A comparative study of crowd solidarity among emergency survivors. *Br. J. Soc. Psychol.* **48**, 487–506 (2009).
- 26. Drury, J. & Reicher, S. Collective Psychological Empowerment as a Model of Social Change: Researching Crowds and Power. *J. Soc. Issues* **65**, 707–725 (2009).
- 27. Guo, D. et al. DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning.

 Preprint at https://doi.org/10.48550/arXiv.2501.12948 (2025).
- 28. Achiam, J. et al. GPT-4 Technical Report. Preprint at https://doi.org/10.48550/arXiv.2303.08774 (2024).
- 29. Neumann, G. & Kawaoka, Y. Highly pathogenic H5N1 avian influenza virus outbreak in cattle: the knowns and unknowns. *Nat. Rev. Microbiol.* **22**, 525–526 (2024).
- 30. Wolff, S. Trump signals he will start pushing for a new world order in first 100 days. *The Conversation* http://theconversation.com/trump-signals-he-will-start-pushing-for-a-new-world-order-in-first-100-days-247594 (2025).
- 31. Forscher, P. S. *et al.* The Benefits, Barriers, and Risks of Big-Team Science. *Perspect. Psychol. Sci.* **18**, 607–623 (2023).
- 32. Chambers, C. D. Registered Reports: A new publishing initiative at Cortex. Cortex 49, 609–610 (2013).
- 33. Yarkoni, T. The generalizability crisis. *Behav. Brain Sci.* **45**, e1 (2022).
- 34. Priem, J., Piwowar, H. & Orr, R. OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts. Preprint at https://doi.org/10.48550/arXiv.2205.01833 (2022).
- 35. McNeish, D. M. & Stapleton, L. M. The Effect of Small Sample Size on Two-Level Model Estimates: A Review and Illustration. *Educ. Psychol. Rev.* **28**, 295–314 (2016).
- 36. Anthropic. Claude 3.7 Sonnet and Claude Code.

https://www.anthropic.com/news/claude-3-7-sonnet (2025).

- 37. Bojić, L. *et al.* Comparing large Language models and human annotators in latent content analysis of sentiment, political leaning, emotional intensity and sarcasm. *Sci. Rep.* **15**, 11477 (2025).
- 38. Thapa, S. *et al.* Large language models (LLM) in computational social science: prospects, current state, and challenges. *Soc. Netw. Anal. Min.* **15**, 4 (2025).
- 39. World Health Organization. The history of zika virus.

 https://www.who.int/news-room/feature-stories/detail/the-history-of-zika-virus (2016).
- 40. Merton, R. K. Priorities in Scientific Discovery: A Chapter in the Sociology of Science. *Am. Sociol. Rev.* **22**, 635–659 (1957).
- 41. Strevens, M. The Role of the Priority Rule in Science. J. Philos. 100, 55–79 (2003).

Data availability

With the exception of Altmetric data, all data are openly available on the Open Science Framework (https://osf.io/a4pby/?view_only=837fdeeeb2eb41a0911fe90858705ee0). Altmetric data cannot be openly shared due to its proprietary nature. However, encrypted copies of Altmetric data are available on the OSF for peer review purposes, and scrambled copies of the data are available for computational reproducibility.

Code availability

All code is openly available on the Open Science Framework https://osf.io/a4pby/?view_only=837fdeeeb2eb41a0911fe90858705ee0.

Acknowledgements

L.W. was supported by the National Science Foundation (grant SOS:DCI 2239418).

Author contributions

Conceptualization: NAC, SF, LW

Data Curation: NAC, JFGBT Formal Analysis: NAC, JFGBT

Funding acquisition: NAC, SF, LW

Investigation: NAC, JFGBT

Methodology: NAC, JFGBT, SF, LW Project administration: NAC, JFGBT Software: NAC, JFGBT, LW

Supervision: NAC, SF, LW

Validation: JFGBT, SF, LW

Visualization: NAC

Writing—original draft: NAC

Writing—review and editing: NAC, JFGBT, SF, LW

Competing interests

The authors declare no competing interests.

Table 1. Research questions, hypotheses, and planned sample, analysis, and interpretations.

Question	Hypothesis	Sampling plan	Analysis plan	Interpretation for different outcomes
During urgent societal events, what is the best characterization of the relationship between the number of co-authors on a scientific article andthe number of days between the event and the publication of the article (as indexed by OpenAlex)?	H1: Increases in team size will be associated with curvilinear returns. Initial increases in team size will be associated with increases in speed, but further increases in team size will be associated with reductions in speed.	A Monte Carlo power simulation using pilot data suggested that 100,000 papers would provide 95% power to detect all hypothesized relationships.	Primary We used mixed-effect regression – weighted by the inverse frequency of observed team size – to separately model linear, logarithmic, and quadratic relationships with team size (included as a continuous variable).	Linear model Positive β_1 : Constant rate of improvement as team size increases Negative β_1 : Constant rate of decline as team size increases Logarithmic model Positive β_1 : Diminishing returns
the number of times the publication is cited by other scholarly articles (as indexed by OpenAlex)?	H2: Increases in team size will be associated with diminishing returns. Initial increases in team size will be associated with increased mentions in scholarly articles, but the magnitude of the advantage will diminish as teams become larger.		E.g., consider speed, operationalized as the number of days since the urgent societal event (e). Linear model: Speed $\sim \beta_{0e} + \beta_1(TeamSize) + \varepsilon$ Logarithmic model: Speed $\sim \beta_{0e} + \beta_1\log(TeamSize) + \varepsilon$	Negative β_1 : Diminishing losses Quadratic model Positive β_1 : Curvilinear returns Negative β_1 : Curvilinear losses
the number of times the publication is mentioned in news outlets (as indexed by Altmetric)?	H3: Increases in team size will be associated with curvilinear returns. Initial increases in team size will be associated with increased mentions in news outlets, but further increases in team size will be associated with reductions in news outlet mentions.		Ouadratic model: Speed $\sim \beta_{0e} + \beta_1(TeamSize) + \beta_2(TeamSize^2) + \varepsilon$ Note $\beta_{0e} = \beta_0 + u_{0e}$ Secondary We will add random slopes to the best fitting model to estimate how relationships vary across urgent societal events (e).	
the number of times the publication is mentioned in public	H4: Increases in team size will be associated with curvilinear		E.g., Mixed linear model	

policy documents (as indexed by Altmetric).	returns. Initial increases in	Speed $\sim \beta_{0e} + \beta_{1e}$ (TeamSize) + ε	
	team size will be associated with	<u>Note</u>	
	increased mentions in policy documents, but further increases in team size will be associated with reductions in policy document mentions.	$oldsymbol{eta}_{1e} = oldsymbol{eta}_1 + u_{1e}$	

Supplementary Information

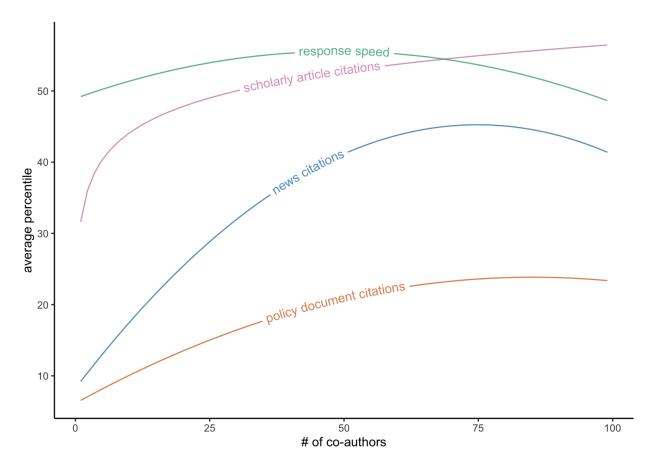
A pilot investigation involved research articles containing keywords linked to three events: (1) the initial release of the generative AI tool, ChatGPT (n = 35,291), (2) the COVID-19 pandemic (n = 1,323,903), and (3) the 2001 World Trade Center attacks (n = 22,723).

Weighted linear regression indicated that each additional co-author on a published research response was associated with a (a) 0.19 percentile increase in scholarly citations (95% CI [0.19, 0.19], t(1382544) = 180.55, p < .001), (b) 0.30 percentile increase in news citations (95% CI [0.30, 0.30], t(539112) = 135.89, p < .001), (c) 0.16 percentile increase in public policy document citations (95% CI [0.16, 0.16], t(539033) = 89.45, p < .001), and (d) .006 percentile decrease in days to publish (95% CI [-0.006, -0.006], t(478544) = -6.91, p < .001). Expressed in raw units, each additional co-author was associated with 1.08 more scholarly citations, 0.38 more news citations, 0.02 more policy citations, and an extremely minor .07 day (2 hour) delay in publication speed.

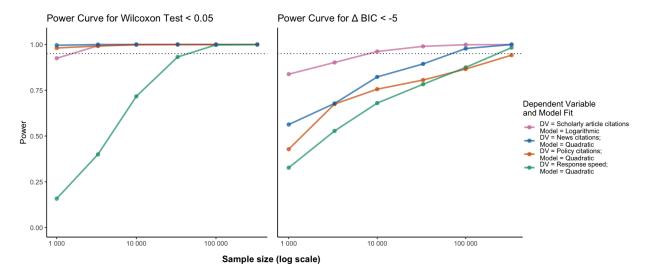
Of course, linear regression assumes that increases in team sizes are associated with *constant* (linear) rates of returns in speed and impact. However, it is unlikely that such relationships are actually linear. For instance, increases in team size may yield *diminishing* returns in terms of impact²³ – building towards a plateau where scientists must make increasingly larger investments to increase the speed and impact of their work. Alternatively, this relationship may be more *curvilinear*: increases in team size may lead to improved performance up to a point, after which further growth becomes *detrimental*. Such possibilities were evaluated via logarithmic and quadratic regression, respectively.

In our pilot analysis, the relationship between team size and scholarly citations was best captured by a logarithmic model, indicating diminishing returns. This model showed substantial improvement over both the linear (Δ BIC= 23,985) and quadratic models (Δ BIC= 5,073), with β_1 = 7.94 (95% CI [7.94, 7.94], t(1382401) = 239.49, p < .001). For all other outcomes, quadratic models achieved best fit. For news citations, the quadratic model improved over linear (Δ BIC= 5,466) and logarithmic models (Δ BIC= 772), with β_1 = 0.99 (95% CI [0.99, 0.99], t(538959) = 104.11, p < .001) and β_2 = -0.01 (95% CI [-0.01, -0.01], t(539076) = -74.33, p < .001). The quadratic model also best fit policy citation data (linear Δ BIC= 1,101; logarithmic Δ BIC= 530), with β_1 = 0.42 (95% CI [0.42, 0.42], t(539029) = 53.59, p < .001) and β_2 = -0.002 (95% CI [-0.002, -0.002], t(1382542) = -81.72, p < .001). Similar patterns were also found for speed (linear Δ BIC= 6,652; logarithmic Δ BIC= 6,081), with β_1 = 0.27 (95% CI [0.27, 0.27], t(691381) = 77.59, p < .001) and β_2 = -0.003 (95% CI [-0.003, -0.003], t(1382542) = -81.72, p < .001). These quadratic effects suggest that while larger teams initially benefit impact and speed, performance often declines beyond a threshold. These estimated points of decline were 75 co-authors for news citations (z = 13.33), 85 for policy citations (z = 15.28), and 49 for speed (z = 8.42). Such sizes lie far outside typical team norms, indicating that the limits of collaboration are rarely exceeded (see Supplementary Figure 1)²¹.

Supplemental Figure 1. Modeled relationships between the number of paper co-authors (x-axis) and four outcomes: response speed (green), mentions in scholarly articles (pink), news media (blue), and policy documents (orange). Analyses are based on 1,382,547 papers published within three years after the release of ChatGPT, the COVID-19 pandemic, and September 11 terrorism attacks.



Supplemental Figure 2. Power curves from Monte Carlo simulations for detecting improved fit of hypothesized models over a baseline linear regression model. Improvements are assessed using non-parametric tests of squared error reduction (Wilcoxon test, left) and differences in Bayesian Information Criteria (Δ BIC, right). The dotted line indicates the 95% power threshold commonly used to determine sufficient sensitivity.



Supplemental Table 1. List of eligible urgent societal events, including their approximate date and associated search terms. For each event, we reported the total number of papers indexed in OpenAlex that mention the keyword, the number of papers published in the 3 years before and after the event, and the estimated growth in publication volume across those periods.

			# of OpenAlex records			
event	approximate date	search term	total	3 years prior to event	3 years after event	Ratio of growth
9/11 Terrorist Attacks	September 11 2001	9/11 terrorism	169100	1393	6969	5.00
2002-2004 SARS Outbreak	November 16 2002	SARS outbreak	160000	224	5855	26.14
Indian Ocean Tsunami	December 26 2004	Indian Ocean Tsunami	8901	6	751	125.17
Hurricane Katrina	August 29 2005	Hurricane Katrina	51940	25	9341	373.64
H5N1 Bird Flu Global Concern	January 13 2006	H5N1	47410	451	1990	4.41
iPhone Launch	January 9 2007	iPhone	104100	65	3994	61.45

Global Financial Crisis	September 15 2008	Global Financial Crisis	703500	27310	58710	2.15
H1N1 Swine Flu Pandemic	April 15 2009	H1N1	120400	2729	25880	9.48
Haiti Earthquake Disaster	January 12 2010	Haiti Earthquake	20380	254	5513	21.70
Deepwater Horizon Oil Spill	April 20 2010	Deepwater Horizon Oil Spill	11420	26	2442	93.92
Arab Spring Begins	December 17 2010	Arab Spring	113000	4042	10240	2.53
Fukushima Nuclear Disaster	March 11 2011	Fukushima Nuclear Disaster	19140	24	4622	192.58
South Sudan Independence and Civil War	July 9 2011	South Sudan Independence	8560	205	1114	5.43
CRISPR-Cas9 Discovery Application	June 28 2012	CRISPR-Cas9	141000	73	2291	31.38
Bangladesh Factory Collapse	April 24 2013	Bangladesh Factory Collapse	5657	363	857	2.36
Ebola West Africa Outbreak	December 26 2013	Ebola outbreak	53050	1136	10710	9.43

Zika Virus Epidemic	April 23 2015	Zika virus	57010	374	19610	52.43
Pope Francis Climate Encyclical	June 18 2015	Pope Francis Climate Encyclical	1811	35	615	17.57
Syrian Refugee Crisis Peak	September 2 2015	Syrian Refugee Crisis	33790	1464	5331	3.64
Flint Lead Poisoning Crisis	September 24 2015	Flint water crisis	5926	383	774	2.02
Paris Climate Agreement	December 12 2015	Paris climate agreement	111000	5755	15730	2.73
LIGO Gravitational Wave Detection	February 11 2016	LIGO gravitational wave	24080	1682	4962	2.95
AlphaGo Beats World Champion	March 15 2016	Alphago	5358	5	1667	333.4
Brexit Referendum	June 23 2016	Brexit	77740	348	28610	82.21
First Commercial Drone Delivery Service	December 7 2016	Commercial Drone Delivery Service	9394	472	1929	4.09
Women's March Following Trump Inauguration	January 21 2017	Women's March Trump	5012	333	1193	3.58

WannaCry Ransomware Attack	May 12 2017	WannaCry	3114	4	1487	371.75
First Gene Therapy FDA Approval	August 30 2017	Kymriah	2078	2	836	418
#MeToo Movement Emergence	October 15 2017	МеТоо	14920	65	4556	70.09
Opioid Crisis Declaration	October 26 2017	Opioid Crisis	34270	3277	10250	3.13
Facebook Cambridge Analytica Scandal	March 17 2018	Cambridge Analytica	7613	82	3118	38.02
Greta Thunberg's Climate Strike	August 20 2018	Greta Thunberg	4163	6	1190	198.33
Indonesia Sulawesi Earthquake and Tsunami	September 28 2018	sulawesi earthquake tsunami	1590	129	517	4.01
IPCC 1.5°C Special Report	October 8 2018	IPCC special report	4862	497	1031	2.07
US Vaping-Related Lung Injury Outbreak	August 1 2019	Vaping lung injury	2924	137	1545	11.28

COVID-19 Pandemic Begins	December 31 2019	COVID 19	2130000	3720	1392000	374.19
George Floyd Murder and Racial Justice Movement	May 25 2020	Black Lives Matter	23380	4043	12430	3.07
Beirut Port Explosion	August 4 2020	Beirut Port Explosion	1469	222	531	2.39
Synthetic Meat Market Introduction	December 2 2020	Cultured meat	2403	324	864	2.67
Bitcoin Mainstreaming	January 3 2021	Cryptocurrency	52260	1079	11590	10.74
Mars Perseverance Rover Landing	February 18 2021	Mars Perseverance	3382	343	1440	4.20
Tonga Volcanic Eruption	January 15 2022	Tonga Volcanic Eruption	3494	213	1207	5.67
Russia-Ukraine War Energy Crisis	February 24 2022	Russia Ukraine War	123900	14180	48270	3.40
Mpox (Monkeypox) Global Outbreak	May 7 2022	Monkeypox	13460	545	8333	15.29
US Supreme Court Dobbs Decision	June 24 2022	Supreme court dobbs	2175	93	1236	13.29

ChatGPT Public Release	November 30 2022	ChatGPT	47480	58	42460	732.07
Silicon Valley Bank Collapse	March 10 2023	Silicon Valley Bank Collapse	3591	265	696	2.63
OpenAl Leadership Crisis	November 17 2023	OpenAl Governance	1267	117	1107	9.46